UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

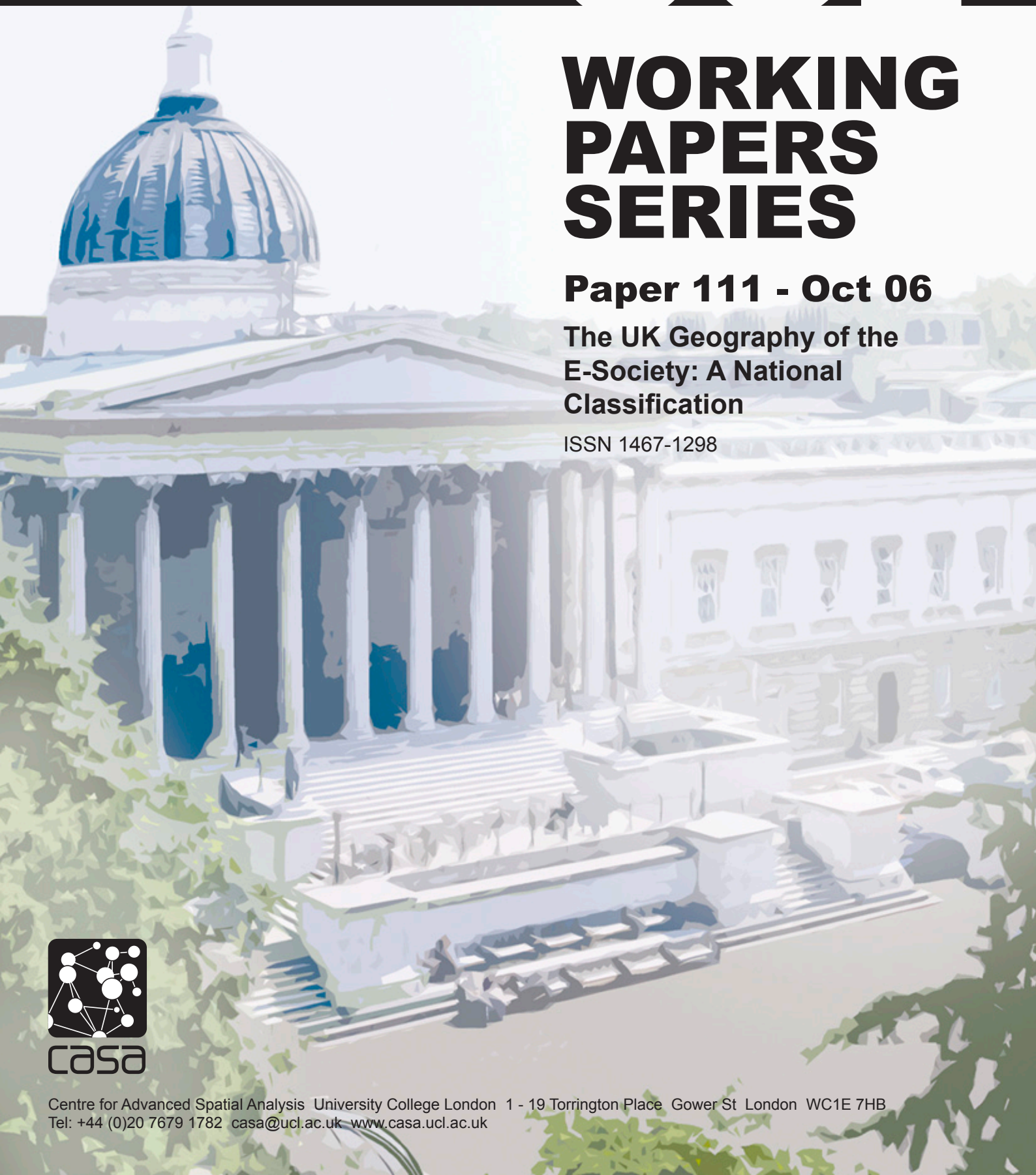
UCL

WORKING PAPERS SERIES

Paper 111 - Oct 06

The UK Geography of the E-Society: A National Classification

ISSN 1467-1298



casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

The UK Geography of the E-Society: A National Classification

Paul Longley*, Richard Webber and Chao Li

Centre for Advanced Spatial Analysis,

University College London,

1-19 Torrington Place, London WC1E 7HB

* Corresponding author. Email: plongley@geog.ucl.ac.uk; Tel: +44 (0) 207 679 1782;

Fax: +44 (0) 207 813 2843

The UK Geography of the E-Society: A National Classification

Abstract

It is simplistic to think of the impacts of new information and communication technologies (ICTs) in terms of a single, or even small number of, 'digital divides'. As developments in what has been termed the ‘e-society’ reach wider and more generalised audiences, so it becomes appropriate to think of digital media as having wider-ranging but differentiated impacts upon consumer transactions, information gathering and citizen participation. This paper describes the development of a detailed, nationwide household classification based on levels of awareness of different ICTs; levels of use of ICTs; and their perceived impacts upon human capital formation and the quality of life. It discusses how geodemographic classification makes it possible to provide context for detailed case studies, and hence identify how policy might best improve both the quality and degree of society’s access to ICTs. The primary focus of the paper is methodological, but it also illustrates how the classification may be used to investigate a range of regional and sub-regional policy issues. This paper illustrates the potential contribution of bespoke classifications to evidence-based policy, and the likely benefits of combining the most appropriate methods, techniques, datasets and practices that are used in the public and private sectors.

Key words

Geodemographic classification, e-society, new information and communication technologies, Great Britain

1. Background: 'E-technology' and disadvantage

In most of the areas of activity in which government is involved, policy makers take express interest in the potential distributional effects of specific programmes or policy options. The most logical scales at which these distributional effects may be considered are at the level of the individual, or household, and at the level of the neighbourhood. Thus some policies, such as the exemption of persons aged over 65 from having to pay for a television licence, are designed to compensate for material inequalities between people of different incomes, or the likelihood that they have earned incomes. Other policies, such as for example those addressing the regeneration of poor neighbourhoods, recognise the extent to which individual and households with multiple forms of disadvantage find themselves disproportionately concentrated in a limited number of neighbourhoods, many of which can be identified using statistical information of the sort that is included in the current Index of Multiple Deprivation (see, for example, Senior 2002; Harris and Longley 2002).

The questions traditionally included in (typically) decennial census questionnaires provide quite a reliable insight into the various forms of disadvantage felt to be most serious by the government. Questions contained in the UK Census of Population which have historically contributed to indicators of neighbourhood deprivation include whether or not members of the household are employed; whether or not they have access to a car; whether or not the dwellings in which they live enjoy a full set of modern amenities; whether or not they live at an acceptable number of rooms per person; whether they are

unable to work on account of long term disability and whether they consider their health to be good or bad. The level of their educational attainment is also considered an important indicator, in this case of employability. These various census data resources are discussed in Rees et al (2002), Dugmore (2004) and Howick (2004).

The set of questions contained in the UK Census includes many that are used in other countries. However there are differences. Censuses in South American countries will typically include information on access to water, to sanitation and to electricity as well as on literacy. Censuses in many countries other than Britain ask direct questions on personal or household income. A very interesting question included in the Australian census of 2000 concerns access to the Internet: respondents were asked whether they had access to a computer at home or at work. They were also asked whether and if so how they accessed the Internet, for instance whether at home, at work or via an Internet café. The rationale for including the Internet in the Australian census was that in an increasingly information based economy, the level of access to and knowledge of how to use personal computers and the Internet is likely to become an increasingly important source of advantage and disadvantage. As information becomes increasingly commodified (Openshaw and Goddard 1987), so the lack of access to a computer, or a lack of knowledge of how to use it, may in time become as significant a source of disadvantage as for example access to or lack of access to a car or access to or lack of access to central heating. Arguably, in terms of employability, a lack of competence in the use of information technology may become as significant a barrier as a lack of higher

educational qualifications. The regional and local dimensions of such inequality remain an important focus for regional policy and the interventions of devolved government.

We can identify at least four sources of disadvantage that might arise from such inequality. First, disadvantage may result from lack of understanding of or access to electronic technologies, and this in turn may manifest itself in various ways. As the Internet becomes increasingly established as *the* medium of choice whereby individuals gain access to human knowledge, an individual without understanding of or access to electronic technologies becomes disadvantaged in terms of human capital. For example, an individual who misses a radio programme which covers an area of specific interest but who is unable to visit the BBC Website in order to read a transcript or to hear a copy of this programme, may find general development of his or her skills and competencies undermined, with implications for the formation of human capital.

Second, lack of understanding of and access to electronic technologies is also a handicap in the labour market. Businesses not unreasonably consider it to be their responsibility only to train new recruits in skills which are specific to their business. A general grounding in information technology is no longer considered a specific skill, and training in general IT skills is considered the responsibility of the employee. If potential employees are expected to be conversant with the basic functions involved in the use of personal computers, those who have not learned them will find it increasingly difficult to obtain any other than unskilled jobs.

A third source of disadvantage that arises from a lack of competence in electronic technologies involves people in their role as consumers. If people are unable to book airline tickets over the Internet it is likely that they will pay higher fares on the telephone, or to book with an airline that sells through agents. Savers unable to access and operate an Internet account are unlikely to obtain as competitive interest rates on their savings as those that do. Increasingly consumers who are unable to access the Internet will find themselves denied access to key information enabling them to make the optimal decisions that informed consumers are expected to make. Lack of access of this information is becoming an increasing source of disadvantage as support which was previously supplied directly by the state, such as education, health care or pensions, is increasingly delivered through agencies that compete on the basis on consumer choice. Consumers, when making decisions regarding issues in the public sector, such as schools for their children, hospitals for their treatment, or opting in or out of state retirement savings schemes, are increasingly expected to use the Internet to find the information on the basis of which such choices can be intelligently made.

A fourth and final source of disadvantage is what one might describe as social exclusion. Schoolchildren who lack access to modern communications technologies are at a disadvantage when wanting to participate in social networks with other pupils in their school. Quite apart from difficulties of communication, the lack of access to the technology can make it more difficult to be accepted by social networks which are based on high levels of peer group communication. Likewise adults without access to electronic technologies find themselves progressively more excluded from the information and the

networks through which social capital is developed and maintained and will, as time goes on, become progressively marginalised from the communities to which they have in the past belonged.

2. A national classification of engagement with the 'E-Society'

The purpose of this paper is to describe the creation of a quantitative model whereby any adult in Great Britain could be evaluated according to their likely level and manner of engagement with electronic technologies. By creating this model we believe that it may be possible to address some of the consequences of the lack of any corresponding question on the 2001 census, namely the ability to estimate variations in the likely level of access to electronic technologies for a set of geographical areas as small as the census output area. We hope that, in so doing, we can actually create something that may be *more* valuable than any single univariate indicator of engagement with technologies – that in practice come in different guises and present different degrees of flexibility in terms of access to and exchange of information. These technologies are also, in practice, made available through rapidly changing pricing structures, and new types of device generally facilitate greater ease of user interaction in the later, mature stages of product lifecycle development. In addition to small area aggregations, we also believe that the database created using the model could and should be used as a device for coding the respondents to field surveys and citizen registers according to their estimated levels and types of engagement. Using a likely measurement of electronic engagement as a field for

analysis, it is possible to better understand the profiles of users of specific services and the extent to which they are likely to have poor or good access to electronic communications and technologies. For example by analysing existing users of 'NHS Direct' (a telephone service run by the UK National Health Service) by these population segments we would be able to understand the extent to which any campaign to extend use of the service would be prejudiced by the lack of ability of low engagement groups to master the necessary use of electronic technologies. Such an analysis might show that the population groups which are most likely to make use of the NHS are ones which tend to have particularly low levels of engagement with these technologies.

As such, this work is predicated upon the belief that the extent to which individuals do or do not engage in electronic communications is not one dimensional. It is our assumption, based on the results of analysis undertaken over a number of years by Experian, one of Britain's leading holders of information on consumer behaviour (www.experian.co.uk), that a more effective way of discriminating between individuals in terms of their manner of engaging with electronic communications is by grouping them into a set of categories which are multivariate rather than one dimensional in their definition. The rationale for this is as follows:

- A significant, but not over-riding reason for variation in the use of electronic communications arises because of variation in household income. At the lower end of the household income distribution we find households that have the desire to engage in electronic communications but not the financial means. There is no direct income measure available in the UK Census of Population.

- Likewise age is also an important discriminator of how people engage with electronic communications. In general, younger age groups are more likely to engage with electronic communications because this enhances their self image, presents them in a good light among their peers or even, at a more basic level, enables them to engage with and communicate with other members of a peer group. They have grown up with the equipment and may have learned to use it at school. Older age groups tend to use electronic communications in a more instrumental manner, in other words as a tool for achieving other benefits rather than as a form of leisure activity or entertainment.
- A third dimension, relevant both to the level of engagement and to the manner of engagement, is the extent to which people are still at the stage of their careers when human capital is being formed. We concur with the view that, insofar as electronic communications provide access to accumulated human knowledge, they will be used more intensively and in a different ways by population groups which are currently eager to progress their career prospects by acquiring a deeper awareness of the knowledge which is a distinguishing feature of their current or proposed profession.

These different dimensions are not only manifestations of differences in means and motivation for engaging with technology, but also guide the selection of technologies and the features of those technologies that users make use of. For instance, for an older self employed tradesman the key benefit of a mobile phone is as a device for picking up messages from existing and future customers. For a mother, the benefit may be in the

form of greater security by being able to keep in touch with her children. For the teenager it may provide a necessary requirement for access to or sustaining membership of a particular peer group.

It is for these reasons that the methodology used in this paper is designed in such a way as to create a set of discrete market segments or behavioural clusters. It is not the purpose of our research to reflect their overall level of engagement through a score on a continuous dimension. It is also for these reasons that we have sought to construct a segmentation of adults at the person rather than the household level. We recognise that within a household there may well be a series of different attitudes and practices among different members of the household. Often differences both in gender and age result in different family members being ideally described by different clusters or market segments.

3. Strategies for grouping individuals on the basis of use of electronic communications

When one seeks to group individuals on the basis of their actual or likely use of electronic communications, there are a number of different approaches which can be adopted (e.g. see Longley 2005). One fairly obvious strategy involves the undertaking of a field survey of a representative sample of adults, asking them a series of questions covering behaviours relevant to the classification and then grouping respondents into

natural clusters on the basis of their response. Such an approach is appropriate if one's interest is primarily in understanding the inter-relationship between various factors influencing that behaviour; or if one is interested in defining market segments that have different needs, that access services in different ways and which are best reached using different communications media. However such a strategy is less effective if one's interest is in obtaining 'external' data with which to enrich the contents of a customer file or a prospect database, or if one wants to create some mechanism for data linkage of different fields contained on different databases.

The principal reasons for these constraints are that any field survey is likely to provide behavioural information on only a small proportion of the total population. Matches between the field survey respondents and names and addresses on other files that one may want to analyse will usually be too infrequent to generate sufficient records for statistically reliable analysis to be undertaken. Such direct data matching is also likely to contravene the provisions of various data protection acts.

One method which researchers have used in an attempt to overcome the limitations of this approach is to evaluate the sampled population in terms of a number of attributes, not necessarily germane to the focus of the study, which can nonetheless be established for significant numbers of the non sampled population. By this means cluster segments can be attributed to large numbers of non-sampled people whose names and addresses as well as other attributes are held on another file (Leventhal 1997). This involves indirect rather than direct fusion. By examining the differences in these other typically demographic

characteristics across the different segments of the study population, it may then be possible to create a form of model whereby non sampled individuals, on the basis of such other information as may be known about them, can be attributed to the best fit cluster which will have been created using information specific to the sampled population. Whilst this approach often seems attractive in theory it typically disappoints in practice. This is because these supplementary information items, even when taken in aggregate, are often a relatively poor predictor of a number of behaviours or product categories – of which we fear that electronic technologies may be one.

An alternative to this approach is to build a form of mathematical scorecard, based on the sample of the population that has been interviewed, and to then apply this formula to the sample of population that has not been interviewed. This approach also has its merits. Typically it tends to provide better discrimination than the cluster and attribution method described above. In situations where one seeks to obtain an ordinal ranking it would generally prove more appropriate than in situations, such as with the study of engagement with electronic technologies, where one presumes that there are qualitatively different forms of e-engagement and where one believes that a one-dimensional ordering may not be appropriate. Statistically a method such as this tends to assume either a linear or some other pre-specified form of non linear relationship between the variables which one has available for modelling and the attributes which one is seeking to predict (but see Wrigley et al 1988). Such a method is also rather difficult to apply in situations where the base data for the population to which one is attributing scores or segment codes are being

continuously refreshed – as with data on the diffusion and use of new information and communication technologies (ICTs: see, for example, Masser et al 1996).

A third method, used in this research, is typically described in the literature as the 'Mosaic – Pixel Grid Methodology' or MPG methodology for short (Webber 2004). This has successfully been used by Experian to build industry specific segmentations at the level of the individual or the household. Industry sectors to which it has been applied include Personal Financial Services (Financial Strategy Segments), Communications (Touchpoint Segments) and Apparel (Fashion Segments). The MPG methodology was conceived by Experian International Ltd. as an efficient method of assigning segmentation codes in a situation where a limited number of attributes were known about a very large population, and in which one wishes to make as accurate as possible assignment for every individual in the country, but where is not essential that every individual assignment should be wholly accurate.

In summary, UK applications of the method make use of two substantial databases. The first of these consists of self completion questionnaires submitted by respondents to 'lifestyle surveys'. These surveys have much wider coverage than conventional market research surveys but at the cost of much less representative response. They invite respondents, often in return for entry in a prize draw, to supply preferences across a wide variety of products and services. The representativeness of such surveys has not been extensively researched (but see Longley and Harris 1999).

The second database contains a set of attributes which is known about virtually all persons whose names and addresses are recorded on the 2003 Electoral Register. The data sources from which these attributes originate are the Post Office's National Postal Address File, the Electoral Register, the file of Company Director's names and addresses which is maintained at Companies House, various shareholder registers and a national classification of residential neighbourhoods, developed at the level of the full postcode and reliant principally on data sources originating from the 1991 Census of Population. We briefly summarise the characteristics of each of these attributes in turn.

The effectiveness of the MPG methodology relies on the ability of Experian's lifestyle survey questionnaires to include questions which are relevant to the study of consumers' engagement with electronic communications and technologies, and on their relevance to the estimation of this behaviour of the limited number of attributes held by Experian about each adult in Great Britain. These attributes, which are collected, updated and maintained by Experian on an annual basis, are derived from public data sources. This fact is of material importance to ensuring that the use of the method does not infringe data protection regulations and that it does not compromise the privacy of individuals and the confidentiality of data relating to them. Here, we briefly summarise the characteristics of each source in turn.

3.1 The Post Office's National Postal Address File

The Post Office's National Postal Address File contains a complete register of all valid postal addresses in Great Britain together with their current postcodes. The purpose of this file is to enable computer bureaux to append correct postal codes to postal addresses. The file is updated quarterly. Using this file, it is possible to categorise individual postal addresses on the basis of a series of text strings that appear in their entries on the file. For example by searching for the word 'farm' in the appropriate field in the postal address it is possible to identify a set of postal addresses which currently are, or which may in the past have been, buildings isolated from towns and villages and set in rural environments. Likewise it is possible to identify addresses with a named house in the first line of the address field, such as Talbot House, Broadlands Road, London N6, and to distinguish these from addresses with a numbered house, such as 16 Broadlands Road. A third useful address form consists of entries which typically denote flats, such as for example Flat 4, 14 Broadlands Road or 22A Broadlands Road.

Although these address forms cannot be guaranteed on their own to identify individuals living in rural locations, in better off outer suburbs or in villages, in conventional urban streets on in divided buildings, when combined with other information they do increase the probability with which one can distinguish population groups which are significantly different in terms of rurality, income and household composition.

3.2 The Electoral Register

The Electoral Register is another important source of information which can be attributed to virtually all individuals and addresses. One of the more obvious data values that can be

derived from the Electoral Register is the likely composition of a household in which any individual elector falls. For example it is reasonable to assume a Richard Webber and a Patricia Webber living at the same postal address can be classified as a married couple. This may not necessarily be the case in every instance but it is likely to be the case more often than not. A Richard Webber living at an address without any evidence of other electors present would be classified as a single male and a Patricia Webber living at an address without any evidence of other electors present would be classified as a single female. Other forms of household that can be distinguished are apparent co-habitees, instances of where a Richard Webber might be found at the same address as a Patricia Smith, and complex households where there are three or more electors present with two or more family names. As with the form of the address, the apparent form of the household is not necessarily a wholly reliable indicator on its own. The important consideration is that, when combined with other information known about the individual, it contributes incremental predictiveness to a multivariate model or classification system.

Because Experian updates its national version of the Electoral Register every year and because it tracks the date of all amendments, it is possible to identify the number of years that any particular elector has been present on the Electoral Register at that address. This has made it possible for each current electoral registration entry to include a ‘length of residency’ indicator, currently up to eighteen years. This indicator does not necessarily indicate the length of time that a person has been living at a particular address but the length of time since reaching the age at which that person is entitled to vote, currently eighteen but in earlier years twenty one.

Given that younger people move house more often that old people and that a person under the age of 36 cannot, by definition, have been on the Electoral Register for more than 18 years, length of residency is a useful input into multivariate models of age. From the Electoral Register it is also possible to construct useful proxies for age with a greater degree of accuracy than many people imagine. The first commercial system of this sort, called Monica, was developed by CACI (www.caci.co.uk) on the basis of the forename of the elector. It is common knowledge that the forenames chosen by parents and given to their children vary over time. In the early Twentieth Century it would have been more common than it now is to name one's daughter Ivy, Gertrude or Doris. Today Jessica, Emma or Olivia are among the most popular names of students at university. Fashions in boys' names also change over time, though somewhat less rapidly than in girl's names. From an analysis of forename and age from a sample of four million respondents to lifestyle surveys it has therefore been possible for Experian to establish the age distribution of people with each of the more common names found among British adults. In order to improve the predictiveness of their age estimation model, Experian takes into account not just the forename of the individual but also the forename of any partner with the same surname at the same address. As a result of this a John Smith living at the same address as an Ivy Smith will be credited with an older age estimate than a John Smith living at the same address as an Olivia Smith.

A third important input to the age prediction system is the number of years that the elector has been present on the Electoral Register at that address. Thus a John Smith co-

resident with an Ivy Smith, both of whom will have been resident at their address for eighteen or more years, will be deemed to be of an older age that a John Smith living alone at an address where he has been resident for only two years or fewer. In addition to length of residence and age, the contents of the Electoral Register can be used to infer gender. Most forenames are assigned as either male or female though there are few which are assigned to the category ‘unknown’. These include names such as Robin and Leslie which can be used by either gender as well as a number of the more obscure forenames of recently arrived minority ethnic groups for which Experian has yet to research the gender.

3.3 Company Directors

A third important source of information on individuals is Companies House. Companies House collects and makes public information on private companies. Part of the information that is collected and made available is information on the names and addresses of the directors of a company. By matching these name and address records against the Electoral Register it is possible to identify those electors who are also company directors. They can also be further distinguished according to the size of the company of which they are directors. This makes it possible to differentiate leaseholders who are be joint directors of a company formed to hold and manage the freehold of a block of flats in which they live from captains of industry who are directors of trading companies with significant annual turnover figures.

3.4 Shareholders

The registers of shareholders in quoted companies are also useful sources of information which can be related back to the national version of the Electoral Register maintained by Experian. These files are not as easy to access and are not as comprehensive in their coverage as the Companies House file or the Electoral Register itself. However they do contain useful information for establishing the likely level of net worth of many individuals.

3.5 Geodemographic classification

Other attributions can be made to individuals based on the demographic characteristics of the postcodes or census output areas in which they live. One such classification is Mosaic, and this research used the Mosaic classification that was built principally upon 1991 Census data. We discuss this in Section 5 below. Further details of how geodemographic classifications in general are available in Harris et al (2005: 147-83) and the specific build characteristics of the Mosaic classification are discussed in Webber (2004).

4. Lifestyle surveys and attribution strategies

Experian is one of a number of commercial organisations that has sought to operate a ‘lifestyle survey’ business. ‘Lifestyle surveys’ as they are known are paper questionnaires which are distributed to consumers on a volume basis and which are designed to generate names and addresses of consumers of particular brands of product categories (see

Longley and Harris 1999). To the extent that respondents provide information on the brands and products that they purchase, they then become ‘qualified prospects’ of value to owners of particular brands as names and addresses that they can mail with specific product propositions. Unlike respondents to traditional market research surveys, respondents to ‘lifestyle’ questionnaires are not stratified in such a way as to provide a representative cross section of the population. Indeed very often the respondent population is biased by the lifestyle companies’ own decision to target questionnaires at types of neighbourhood containing consumers of most interest to their prospective clients. On the other hand the volume of responses generated by lifestyle surveys does provide much larger samples when broken down by subgroups such as people living in particular types of neighbourhood or belonging to specific permutations of age and occupation.

The contents of the Experian lifestyle survey were made available to us for the purposes of this research. This represents some 4,000,000 respondents, information having been collected over a two year period. The Experian survey provides information on a wide range of electronic technologies and communications. These items are listed in Table 1.

[Table 1 about here]

In many circumstances such as this it would seem logical, when trying to predict the level of engagement of an individual with electronic technologies and communications, to append these individual level attributes to the results of a survey questioning individuals

on their behaviours in this particular field of study. One would then build a series of models, one for each behaviour, using some form of multiple regression, so that the likelihood of, for example, having access to the Internet at home, is established using a formula based on the relationship between a series of attributes that can be known both about the sample of survey respondents and the rest of the adult population. Though such an approach might be appropriate where the universe of individuals to whom the prediction is applied is relatively small and where the number of attributes known about each individual is large, it was not thought to be a particular efficient approach in this particular situation. This is because our number of predictors is small, these predictors are categorical rather than continuous in form, and both the sample of records for which survey data are available and the universe of records to which the attribution model is to be applied is large.

The attribution process allows use of only nine distinct variables, each with a limited number of classes. Overall, when we multiply together the number of classes of each of the nine variables, we find that the total number of permutations comes to 6,240. For sake of convenience we refer to these permutations as ‘Pixels’. Since the number of possible values generated by a multiple regression formula cannot logically exceed the maximum possible number of permutations of class values on its input variables, it is evident that such a model, when applied to a file of over 35 million electors, could not produce more than 6,240 different estimated values. In other words, across the elector register as a whole, there would be on average some 5,700 individual electors who, because they held identical values on the input variables, would therefore hold identical values output by

the model. It would therefore be much more logical not to attribute scores directly to all individual electors but instead to attribute scores to the 6,240 possible permutations of the input variables, into one of which each individual elector must necessarily fall.

The second implication of there being only 6,240 different possible permutations of classes of the input variables to a model is that one may be able to measure the level of take up of electronic technologies and communications in these permutations directly. Given that we have access to the results of around four million lifestyle survey questionnaires, a Pixel of average size, say, of 650 respondents would provide a sufficiently large sample to allow behaviour to be analysed directly by Pixel rather than indirectly via a mathematical model.

Clearly if information, on for example the ordering of wine through the Internet, is known for four million respondents and if there are 5,700 respondents on average in each permutation, then it will be possible to calculate directly the proportion of the 5,700 respondents in a typical cell who claim to use the Internet to order wine. This is likely to be a sufficient sample in terms of absolute size for the purpose. The average value for the permutation can then be attributed to all non sampled occurrences on the Electoral Register on the basis of their Pixel alone and without recourse to any form of regression model. Using this approach, it may then be possible to dispense altogether with a model, whether built using regression or otherwise, for attributing observations based on the sample of respondents to the wider universe of electors.

In practice the four million lifestyle respondents to whose behavioural information we have access are not distributed at all evenly across the 6,240 permutations. This is not principally because the lifestyle survey respondents form a biased population when compared with the universe of electors, although this is certainly the case. The distribution of electors by 'Pixel' categories is itself very unequal in size, as shown in Figure 1. This skew in the distribution originates from two quite different factors. The first of these is that the population size of some of the classes of the base variables from which Pixel is constructed is very uneven. For example the proportion of persons of unclassified gender is very much smaller than the proportion of persons who are male or female. Likewise the proportion of people who are directors of large companies is very much smaller than the proportion of people who are not directors of any company.

[Figure 1 about here]

The second reason for the skew in the distribution of electors by 'Pixel' category is that many of the classes of the different variables are themselves correlated with each other. Thus it is unlikely that one would find as many as the average number of electors (5,700) who simultaneously met the criteria that they lived in a farm, that they were aged under 24, that they had been on the electoral roll at that address for more than 12 years and that they were of unclassified gender.

Given the nature of the questions on the lifestyle survey, such as 'do you have access to a personal computer' and the fact that none of them were questions where across the entire

universe of respondents fewer than five per cent were likely to tick the ‘yes’ box, we set a cut off for each Pixel that it would require a minimum of 200 respondents in it in order for information relating to its members to be considered statistically reliable. Given a 5% penetration of a product across the survey file this would represent on average ten positive respondents for that question in that Pixel. (Obviously where the penetration is higher than 5% then one would have access to correspondingly more than ten positive respondents in that Pixel).

We therefore found ourselves in the position of splitting the 6,240 Pixels into three groupings, each of which would be subject to a different estimation strategy. For those Pixels with more than 200 survey respondents we would be in a position to use the observed level of market penetration of a product in that Pixel in order to attribute the likely level of penetration of that product among non respondents to the lifestyle survey in those Pixel categories. For those Pixels with no survey respondents and with no occurrences on the Electoral Register, no imputation method is necessary because there would be no non respondents in those Pixels. For the remaining Pixels, that is those with between one and 200 lifestyle survey respondents (or with no respondents but some cases among the non respondent file), we used an imputation method based on multiple regression.

For this final band we therefore calculated a likely level of market penetration for each of 58 behaviours covered in the lifestyle survey using a regression formula. This formula took into consideration the class values of that Pixel type on each of the nine variables

contributing to the Pixel classification. The dependent variables in these 58 regression models were the average penetration rates of these behaviours in each of the Pixel types with more than 200 respondents to the lifestyle survey. The independent variables used were the nine variables the permutations of whose classes make up Pixel. In these regression models population weights were applied to each of the observations in order to compensate for the very uneven population size of the Pixel categories. Thus the observed values for the more populous Pixel categories were given correspondingly greater influence in the optimisation of the regression formula than were the observed values for the less populous Pixel categories.

By this means we have been able to create a table containing one row for each Pixel category represented on the Electoral Register where each of the 58 columns contains the actual or estimated proportion of lifestyle respondents with positive responses to one of the 58 behavioural questions relevant to the field of study. For the Pixel codes containing more than 200 lifestyle respondents the values are real, observed results; for the Pixel codes with fewer than 200 lifestyle respondents the values are estimates, based on the regression models. Though the proportion of Pixels for which these values are estimates is quite high, the proportion of lifestyle respondents represented by these Pixels is very small. As a result the great majority of non respondents to whom behavioural attributions are made will be assigned values based on the actual observed averages rather than on the basis of estimates created by the regression formulae.

5. Incorporating geodemographics

The 6,240 Pixel codes contain the full set of external information which can be known about each British elector based on publicly available information held at the person level. However, in order to better discriminate between individuals according to their likely level of engagement with electronic technologies and products it is appropriate to supplement this person level information with additional summary statistics which describe the characteristics of the residential neighbourhood in which the individual lives. The most efficient means of including neighbourhood as well as individual level attributes into the segmentation process is to make use of one or other of the various geodemographic systems which have been constructed by commercial organisation such as Experian or CACI. These systems seek to classify individual postcodes into a limited number of types of residential neighbourhood which are broadly similar in their scores across a wide range of different statistical indicators. These measures are sourced both from the most recent census and from other public data sources, including those used to originate information at the person or household level. The current Mosaic classification, for example, creates 61 categories (‘Types’) of neighbourhood in this manner, ranging from type A01 ‘Global Connections’ to type K61 ‘Upland Hill Farms’. However, at the start of this research, only a previous version of Mosaic was available. This classification was created as a result of the analysis of the 1991 census statistics and has a set of 52 Mosaic Types.

The information sources used to construct Mosaic provide summary statistics for many different levels of geographical resolution. These include the full postcode, the census enumeration district or output area, the postcode sector and various custom area definitions based on radius around the postcode. However the classification code is attributed at the unit postcode level and the classification is accessed using a correspondence table which gives the classification code for each individual postcode. The current Mosaic classification covers the entirety of the United Kingdom, although the version of Mosaic used in this study covers Great Britain only.

Using the postcode field in the name and address element of the lifestyle survey respondent questionnaire it is clearly possible to analyse the relationship between the behavioural information relevant to electronic technologies and communications and type of residential neighbourhood. With around four million lifestyle respondents there is more than sufficient sample size to support analysis at the full level of 52 Mosaic Types. However the number of lifestyle respondents is not sufficiently large as to generate an adequate sample of respondents from each permutation of Pixel and Mosaic. Some other method is therefore needed to take advantage of the inherent predictive value of type of neighbourhood in assessing propensity to engage in behaviours relevant to our field of study.

The method which we have adopted involved the creation and population of what we refer to as a Mosaic / Pixel grid. This grid contains a cell for each permutation of Pixel and Mosaic and thus contains in the region 300,000 cells, on average one for

approximately thirteen lifestyle respondents. The manner in which these cells are populated with estimated values on the relevant 58 behavioural values from the lifestyle survey is as follows:

- First, using the lifestyle respondent file, we calculate the observed numbers of respondents in each combination of Mosaic Type and Pixel category.
- Next we calculate the average score on the relevant behavioural variables for each Mosaic Type and for each Pixel category. These calculations then form column and row averages for the Mosaic / Pixel grid. We do not calculate these values for every intersection of rows and columns, only the total population on the lifestyle survey.
- Next we compare the average score on each of these 58 variables in each of the 52 Mosaic Types with the level that would have obtained had each respondent within that Mosaic Type behaved in a manner identical to the other respondents in the same Pixel category. Differences between actual and estimated behaviour indicate situations in which the type of neighbourhood in which a person lives has significant incremental effect on his or her behaviour. The Mosaic Type ‘A03: Corporate Chieftains’, for example, could be expected to contain disproportionate numbers of electors who are company directors, who are large shareholders and who lived in named houses. The Mosaic Type ‘J51: Sepia Memories’ by contrast would contain disproportionate numbers of electors who fall into Pixel categories characterised by older age groups and living in address forms characterised by flats. However if, even despite this, the average level of behaviour on particular variables is even higher among Corporate Chieftains than one would expect on

the basis of their Pixel mix then it is safe to conclude that this type of neighbourhood contributes incremental discrimination.

- For each Mosaic Type the 'expected' proportion of lifestyle respondents reporting particular behaviours is compared with the actual observed level, using Mosaic on its own. In the example of the Mosaic Type 'A03: Corporate Chieftains' it would not be surprising if the observed proportions of respondents reporting ownership of a digital camera would be greater than the proportion predicted on the basis of the Pixel mix of that Mosaic Type.
- Values are then attributed to each cell in the Mosaic / Pixel grid using the product of two input values only. The first of these is the average level of the reported behaviour for that Pixel type nationally, across all Mosaic Types. The second is an adjustment based on the extent to which the observed data for the particular Mosaic Type lies above or below the predicted value based on its Pixel mix. For example the observed level of ownership of a digital camera in a certain Pixel category may be 12.3%. The ratio of observed ownership of a digital camera to the level estimated on the basis of its Pixel mix for 'A03: Corporate Chieftains' may be 1.10. In such a situation the projected ownership level of digital cameras in the permutation of a given Pixel type and the Mosaic Type Corporate Chieftains would be (12.3 x 1.10) or 13.5%.

This method was applied to all 58 of the behaviours considered relevant to the analysis of the adoption of electronic technologies that are covered by the lifestyle survey. The

output of this process is a three dimensional database, 6,240 Pixels x 52 Mosaic Types x 58 behaviours.

Once levels of behaviour are attributed to each of the combinations of Pixel and Mosaic, as represented by cells in the Mosaic Pixel Grid, it becomes possible to consider the reduction of this very large classification into a more manageable and more meaningful set of categories. In order to undertake this form of data reduction we use a cluster analysis package which groups together those cells within the Mosaic Pixel Grid that have broadly similar scores across the range of 58 behaviours whose values we have modelled.

The cluster analysis algorithm used to create this typology is one based on a process of iterative location which allocates records to clusters on the basis of least sum of squares. However, as a result of the very unequal population size of the Pixel / Mosaic combinations and in order to achieve a set of categories roughly uniform in terms of population size, the algorithm uses a set of record weights, these being allocated so as to be proportionate to the population size of each record. The effect of this is that the average profile of each cluster is calculated as the populated weighted average profile of the Pixel Mosaic combinations assigned to that cluster, not the pure arithmetic averages. The population weighting also leads to the resulting clusters being more even in terms of population size than they are in terms of their number of Mosaic / Pixel grid cells.

The cluster analysis was applied to the 300,000 Mosaic / Pixel combinations in order to produce a set of 23 clusters which optimally differentiated the original records in such a way as to retain the highest possible variation across the 58 chosen behaviours. Partly on account of the very small size of many of the input records and partly on account of the fact that all the records represented estimated rather than actual true observations, the loss of variance involved in this grouping process turned out to be relatively small (25.74%). Many of the Mosaic Types and many of the Pixel categories could be found almost exclusively in just one cluster. The key benefit of the cluster analysis was that it resolved conflicts in the minority of instances where the Mosaic Type and the Pixel category suggested contrasting likelihoods of behaviour across large number of the lifestyle response categories.

6. The classification and its application.

The outcome of the clustering process is illustrated in the 'tree diagram' shown in Figure 1. Webber (2004) provides a discussion of this visual classification method and how it makes it possible to show the relationships between the Types and higher aggregation Groups that make up a classification. It is possible to describe the characteristics of the classes using a wide range of quantitatively measured behaviours (from the Experian lifestyles survey) as well as conventional uni- or multi-variate social indicators. An additional characteristic of many geodemographic classifications is the ascription of 'pen profiles' to types within any given classification, based upon the constellation of type characteristics (see Harris et al 2005: 171-3; Birkin 1995). Such profiling entails a degree

of subjectivity, but is widely used in order to render classifications intelligible to end users. The textual interpretation of our E-Society classification is presented in Table 2. Together, Figure 1 and Table 2 describe the outcome of assigning each individual to one of 23 types, and how these in turn are nested into eight larger groupings.

[Figure 2 about here]

[Table 2 about here]

The weighted population sizes of each of the types are shown in Table 3. The individual basis to the classification makes it possible to cross tabulate our classification with any population characteristic for which individual characteristics are known – Table 3, for example, presents the gender breakdown of each type, while Table 4 presents age distributions, both derived using the commonly used assumptions inherent in the Experian data.

[Table 3 about here]

[Table 4 about here]

In the sequel to this paper, we will investigate how this classification may be used to develop scenarios for a number of applications in regional and urban policy analysis, such as voter interest in devolved regional assemblies. For present purposes it is sufficient to note the inherent superiority, all other things being equal, of a classification that operates at the individual (adult) level. The classification can be appended to any

social survey that has basic address information, and provides valuable context to secondary analysis of such data (e.g. see Ashby et al 2005). This level of disaggregation also makes the classification a useful device for devising sampling schemes at the regional, national or local level.

Bespoke classifications of this nature have been available to commercial users for some time, yet academic users have often felt ‘locked out’ of such developments by the terms and costs of commercial licences. Our own classification is available to academics for bone fide research purposes through a pioneering data sharing agreement with Experian Ltd. This makes it possible for users to access the data through the UK Economic and Social Research Council (ESRC) Data Archive and through data services provided direct by Experian.

Taken together, we suggest that this classification represents an important development in UK social science data policy and analysis. First, our partnership arrangement with Experian has enabled us to incorporate pertinent, timely disaggregate data on individual behaviours into a nationwide classification. This represents an important development in data policy that is consistent with developments in the practice of social science in the Twenty First Century, and is a partial realisation of one of the central recommendations of the Commission on the Social Sciences (2003). Second, the assembly of data from very different sources and different domains exemplifies the benefits of ‘pooling’ data in order to create bespoke classifications. As such, it exemplifies the ways in which users have become empowered to create bespoke classifications where no generalised

distribution of behaviours is known. The methodology described here is as much informed by solution-centred commercial practices as conventional social survey research practice, and is consistent with multi-sector, interdisciplinary team working to devise solutions to real world problems in real time. Third, in substantive terms, we suggest that our classification establishes a geographically extensive, generalised basis that helps us to discern trends in the future differentiation of British society. In a sequel to this paper, we will explore the nature of the relationships between the groups and types that we have identified and the transition probabilities between them.

[Figure 3 about here]

We believe that our classification holds considerable power. To anticipate some of our arguments, we present an illustrative comparison of the most (Group 1) and least (Group 8) engaged of our **groups**. These groups are quite different in their usage of ICTs (see Table 5). At the regional scale (Figure 3), it is not difficult to identify the correspondence between the geography of e-engagement and aspects of historic industrial structure, such as the distribution of coalfield industries. Such maps highlight the challenges that may face regional development agencies, for example, if potential employers require that technological literacy is a sufficiently ingrained characteristic of regional and local economies.

[Table 5 about here]

[Figure 4 about here]

At the sub regional scale the distributions suggest some possible patterns of disadvantage that might arise from planning initiatives: for example, the Inner London congestion charge is most easily paid using phone text or Internet channels, and access to and familiarity with such technologies may present as big an obstacle to some users as the charge itself. The maps shown in Figure 4 illustrate the likely geographies of these constraints on access.

We will examine these and other aspects of our classification in our future research. Taken together, we believe that this work establishes that geographic and demographic variables are essential in the interpretation of new digital divides and in the profiling of many of the advantages and disadvantages that will characterise the e-society, and provide context to more detailed studies of forms and modes of engagement with new technologies.

7. Acknowledgement

This research was funded under Economic and Social Research Council grant RES-335-25-0020, ‘Digital differentiation: consumption profiles of fracturing digital divides’.

8. References

Ashby D I, Irving B L, Longley P A 2005 Police reform and the new public management paradigm: matching technology to the rhetoric. At **Environment and Planning C: Government and Policy**.

Birkin M 1995 Customer targeting, geodemographics and lifestyle approaches. In Longley P A, Clarke G P (eds) **GID for business and service planning**. Cambridge, GeoInformation International: 104-49

Commission on the Social Sciences 2003. **Great Expectations: the Social Sciences in Britain**. Academy of Learned Societies in the Social Sciences, London

Dugmore K 2004 Why the census is so important to commercial companies. In Dugmore K, Moy C (eds) **A guide to the 2001 Census: essential information for gaining business advantage**. London, TSO (The Stationery Office):13-16

Harris R, Sleight P, Webber R 2005 **Geodemographics, GIS and neighbourhood targeting**. Chichester, Wiley

Harris R J, Longley P A 2002 Creating small area measures of urban deprivation. **Environment and Planning A**, 34: 1073-93

Howick R 2004 A guide to all the statistical outputs. In Dugmore K, Moy C (eds) **A guide to the 2001 Census: essential information for gaining business advantage**. London, TSO (The Stationery Office): 47-58

Leventhal B 1997 An approach to fusing market research with database marketing. **Journal of the Market Research Society** 39(4): 545-88

Longley P A 2005 Urban studies. In K Kempf-Leonard (ed) **Encyclopaedia of Social Measurement**. Elsevier, San Diego: 921-6.

Longley P A, Harris R J 1999 Towards a new digital data infrastructure for urban analysis and modelling. **Environment and Planning B** 26: 855-78

Masser I, Campbell H, Craglia M 1996 **GIS diffusion: the adoption and use of GIS in local government in Europe**. London, Taylor and Francis.

Openshaw S, Goddard J 1987 Some implications of the commodification of information and the emerging information economy for applied geographical analysis in the United Kingdom. **Environment and Planning A**, 19: 1423-39.

Rees P, Martin D, Williamson P 2002 Census data resources in the United Kingdom. In Rees P, Martin D, Williamson P (eds) **The Census data system**. Wiley, Chichester: 1-24

Senior M 2002 Deprivation indicators In Rees P, Martin D, Williamson P (eds) **The Census data system**. Wiley, Chichester: 123-38

Webber R 2004. Designing geodemographic classifications to meet contemporary business needs. **Journal of Interactive Marketing** 5(3), January/March

Wrigley N, Longley P A, Dunn R 1988 Some recent developments in the specification estimation and testing of discrete choice models. In RG Golledge, H Timmermans (eds) **Behavioural Modelling in Geography and Planning**. London, Croom Helm: 96-123

List of figures and tables

Table 1: Variables extracted from Experian Lifestyles Survey in order to classify consumers

Variable Reference	Variable	Description
1	Cable TV - consider	LEISURE - CONSIDER SUBSCRIBING TO CABLE TV
2	Cable TV - have	LEISURE - HAVE CABLE TV
3	Buy mobile - consider	YOUR HOME - CONSIDER BUYING MOBILE PHONE
4	Buy mobile - have	YOUR HOME - HAVE MOBILE PHONE
5	Mobile - company	YOUR HOME – HAVE COMPANY MOBILE PHONE
6	Mobile - personal	YOUR HOME – HAVE PERSONAL MOBILE PHONE
7	Mag.Computing/IT-sub	LEISURE-MAGAZINES – SUBSCRIBE TO COMPUTING/IT TITLES
8	PCuse-modem cons	YOUR HOME – CONSIDER GETTING PERSONAL COMPUTER MODEM
9	PCuse-modem have	YOUR HOME – HAVE PERSONAL COMPUTER MODEM
10	PCuse-PCsys have	YOUR HOME – HAVE PERSONAL COMPUTER SYSTEM
11	Purch compgame mail/	HOME SHOPPING - HAVE PURCHASED COMPUTER GAMES BY MAIL/TELEPHONE
12	Bk holiday-Internet	HOLIDAYS & TRAVEL - BOOK HOLIDAYS VIA INTERNET
13	PC use - PC sys cons	YOUR HOME – CONSIDER GETTING PERSONAL COMPUTER SYSTEM
14	Email at home	YOUR HOME - EMAIL AT HOME
15	Email at work	YOUR HOME - EMAIL AT WORK
16	Purch bks-Internet	HOME SHOPPING - HAVE PURCHASED BOOKS (GENERAL) USING INTERNET
17	Purch childclth-Internet	HOME SHOPPING - HAVE PURCHASED CHILDRENS CLOTHES USING INTERNET
18	Purch compgame-Internet	HOME SHOPPING - HAVE PURCHASED COMPUTER GAMES USING INTERNET
19	Purch fash.wear-Internet	HOME SHOPPING - HAVE PURCHASED FASHION WEAR USING INTERNET
20	Purch music-Internet	HOME SHOPPING - HAVE PURCHASED MUSIC (MEDIA) USING INTERNET
21	Purch plant.etc-Internet	HOME SHOPPING - HAVE PURCHASED SEEDS/PLANTS/BULBS USING INTERNET
22	Purch video-Internet	HOME SHOPPING - HAVE PURCHASED VIDEOS USING INTERNET
23	Purch vitamins-Internet	HOME SHOPPING - HAVE PURCHASED VITAMINS/HEALTH SUPPLEMENTS USING INTERNET
24	Purch wines-Internet	HOME SHOPPING - HAVE PURCHASED WINES USING INTERNET
25	Read ComputingIT mag	LEISURE – MAGAZINES – READ COMPUTING/IT TITLES
26	Bt gds/serv.Net 2-3t	HOME SHOPPING - BOUGHT GOODS/SERVICES 2-3 TIMES USING INTERNET
27	Bt gds/serv.Net 4+t	HOME SHOPPING - BOUGHT GOODS/SERVICES 4+ TIMES USING INTERNET
28	Bt gds/serv.Net Never	HOME SHOPPING – NEVER BOUGHT GOODS/SERVICES USING INTERNET
29	Bt gds/serv.Net Once	HOME SHOPPING - BOUGHT GOODS/SERVICES ONCE USING INTERNET
30	Email is other	YOUR HOME-EMAIL PROVIDER
31	Shop.-Lrn frmNet-S/D	SHOPPING – DO YOU LEARN FROM INTERNET/WEB? – SLIGHTLY DISAGREE
32	Shop.-Lrn frmNet-D	SHOPPING – DO YOU LEARN FROM INTERNET/WEB? - DISAGREE
33	Shop.-Lrn frmNet - Neither	SHOPPING – DO YOU LEARN FROM INTERNET/WEB? - NEITHER AGREE NOR DISAGREE
34	Shop.-Lrn frmNet -A	SHOPPING – DO YOU LEARN FROM INTERNET/WEB? - AGREE
35	Shop.-Lrn frmNet-S/A	SHOPPING – DO YOU LEARN FROM INTERNET/WEB? – SLIGHTLY AGREE
36	NetAccess frm TV-have	LEISURE – HAVE INTERNET ACCESS FROM TV
37	NetAccess frm TV-con	LEISURE – WOULD CONSIDER INTERNET ACCESS FROM TV
38	MobileP NetAccess-have	YOUR HOME - MOBILE PHONE HAS INTERNET ACCESS
39	MobileP NetAccess-cons	YOUR HOME – WOULD CONSIDER MOBILE PHONE WITH INTERNET ACCESS
40	Spend gds/s Net<L10	HOME SHOPPING – SPENDING ON INTERNET GOODS/SERVICES UNDER £10
41	Spendgds/s NetL10-99	HOME SHOPPING – SPENDING ON INTERNET GOODS/SERVICES £10-£99

42	Spd gds/sNetL100-500	HOME SHOPPING – SPENDING ON INTERNET GOODS/SERVICES £100-£500
43	Spd gds/s Net>L500	HOME SHOPPING – SPENDING ON INTERNET GOODS/SERVICES OVER £500
44	WhyDoShop-Net shopping	HIGH STREET SHOPPING – HEAVY INTERNET SHOPPING
45	ShopOnNet-mainGrocer	HIGH STREET SHOPPING - SHOP VIA INTERNET FOR MAIN GROCERY SHOPPING
46	ShopOnNet-othGrocery	HIGH STREET SHOPPING - SHOP VIA INTERNET FOR OTHER GROCERY SHOPPING
47	RecentChanged mobile	YOUR HOME - RECENTLY CHANGED MOBILE PHONE
48	RecentChangMobi-cons	YOUR HOME – RECENTLY CONSIDERED CHANGING MOBILE PHONE
49	HrsSpdperWkonNet<1hr	YOUR HOME - LESS THAN 1 HOUR SPENT PER WEEK ON INTERNET
50	HrsSpdperWkonNet2-5h	YOUR HOME – 2-5 HOURS SPENT PER WEEK ON INTERNET
51	HrsSpdperWkonNet6-10	YOUR HOME – 6-10 HOURS SPENT PER WEEK ON INTERNET
52	HrsSpdperWkonNet10+h	YOUR HOME – 10+ HOURS SPENT PER WEEK ON INTERNET
53	BankFreqUse-Net/PC	MONEY & INVESTMENT - FREQUENTLY USE INTERNET/PC BANKING
54	BuyFinanProdonNet	MONEY & INVESTMENT - BUY FINANCIAL PRODUCTS OVER THE INTERNET
55	BuyFinanProdonNet-co	MONEY & INVESTMENT – WOULD CONSIDER BUYING FINANCIAL PRODUCTS OVER INTERNET
56	BuyFinanProdonNet-No	MONEY & INVESTMENT – WOULD NOT CONSIDER BUYING FINANCIAL PRODUCTS OVER THE INTERNET
57	PurFina/InsProd Net	HOME SHOPPING - HAVE PURCHASED FINANCIAL/INSURANCE PRODUCT USING INTERNET
58	Pur-PC/Per/Softw Net	HOME SHOPPING - HAVE PURCHASED PC/COMPUTER PERIPHERALS/SOFTWARE USING INTERNET

Table 2: The E-Society classification

Group A : E-unengaged

The '*E – unengaged*' are typically groups that do not have access to electronic communications or technologies. Most are too old, too poor or too poorly educated to be able to access them, and instead traditionally rely upon personal contacts they trust for advice. Within this group there are low levels of literacy and many people do not feel that their life outcomes are much subject to their own decisions. Within this group there is a very low level of ownership of personal computers, very little access to them at work and little ambition to master the skills necessary to take advantage of information technologies. Unsurprisingly, these people have a very low level of using email at any location (home, work and other locations) or participating in other on-line activities.

Members of this group tend to live in the poorer areas of traditional mining and manufacturing towns and to have conservative social attitudes. A high proportion of the group is made up of elderly people, many of whom live in social housing or sheltered accommodation.

Type A01 : Low technologists
This type contains a number of people, mostly older women it would seem, whose primary use of the Internet, if they use it at all, is to buy apparel, children's clothes and vitamins. For these people the Internet is seen as an electronic version of a mail order catalogue, and not something that you learn from. Its members are particularly unlikely to own a mobile phone or to subscribe to cable television.

Type A02 : Cable suffices
This type comprises people with some limited interest in electronic technologies but who have neither the education nor income to become heavily engaged in using them. Many of this type are men who have recently retired or who are approaching retirement. A high proportion has access to cable television.

Type A03 : Technology as fantasy
This type contains many old males, some of whom have an interest in electronic technology and like to read about it, but few of whom use it for obtaining information or for on line ordering. This is a group which has very low take up of cable television. Many transient people fall into this category.

Type A04 : Mobile's the limit
This type has particularly low levels of use of computers and the Internet, knows next to nothing about the technology and has no motivation to do so. They enjoy more traditional modes of communication, but the mobile phone represents the limit of their technical ambition. Many of this type are female and elderly.

Type A05 : Too old to be bothered

This type consists mostly of very old people who feel that they predate anything to do with electronic technologies. Members are particularly unlikely to be found purchasing or reading 'techie' magazines and are among the least likely to find the computer a useful medium for playing computer games – or even watching videos.
Members of this type have little interest in acquiring E-technology skills.

Type A06 : Elderly marginalised
This type consists mostly of very elderly adults, many living on their own,, who have very poor levels of access to electronic technology. Technology seems to be moving on at a rate faster than they can keep up with – for this type, mobile phones and cable television are still novelties, never mind personal computers and the Internet.

Group B : E-marginalised

The '*E – marginalised*' are not necessarily averse to the use of electronic technologies but often lack the disposable income to equip themselves with them, or the training and education needed to understand how to make effective use of them. In this group we find very low level of PC ownership and very little use of the Internet to obtain information or to undertake transactions. However there are members of this group who regularly use personal computers to keep in touch via email and more are considering getting on line. This group does use simpler and less expensive technologies such as mobile phones.

Many members of this group are relatively unskilled young workers, many of whom are in manual occupations. Many also live in low rise council estates, in areas of high unemployment, low incomes and where people are reliant upon public services.

Type B07 : The Net ; What's that?
This type has a low level of engagement with electronic technologies. However those that are not engaged have very little interest in acquiring access to personal computers or to the Internet, although they are interested in getting access to a mobile phone. This type contains a large number of people in later middle age.

Type B08 : Mobile Explorers
This type contains many young people. They have a high level of access to the Internet both at home at work. They enjoy using computers to play games and to watch videos but do not use them to acquire information or to undertake transactions. Many of this group are young. They earn ready money and spend a significant amount of it on their mobile phones.

Type B09 : Cable TV heartland

This type lives and works among a peer group for whom technology is an important lifestyle statement. Members invest considerable time considering the purchase of new technologies. They were amongst the earliest adopters of devices that link mobile telephony with the Internet. They read a lot about technology in magazines and spend a lot of time on the Internet. They send a lot of emails but do not make a lot of on-line purchases.

Group C : Becoming engaged

Members of this group often acquire their competence in the use of information technology at work, since many of them are young people working in junior white collar occupations in modern offices. They are keen to become more expert in the use of new technologies and to use them for new applications. Many spend time browsing the Internet but without necessarily making many transactions.

Many members of this group work in large cities and may be starting a life in a house that they own, typically in one of the cheaper inner suburbs. Their use of the Internet at work may be a practice that their employers may be keen to control or reduce.

Type C10 : E-bookers and communicators

This type is a particularly active user of email, receiving and sending messages both at work and while on the move. The type includes a large number of young, single people, who are particularly interested in the media of communications – they are heavy users of mobile phones but also frequent switchers to and adopters of new mobile technologies. Although ownership rates of personal computers are only average, many individuals use computers to order music and fashion on line. Downloading of music is a particularly common activity. But this type does not make use of the latest technical features of information technology and is unlikely to have professional involvement in the IT industry.

Type C11 : Peer group adopters

This type exists on lower income and is younger than its peers in Type C10, and is even more reliant upon email, text messaging and the use of mobiles to participate in peer group activities. Fewer members of this group are employed in the types of job which would allow access to email at work, and many fewer use personal computers to purchase goods on-line. Members of this type are more likely than those in type C10 to have access to cable television and to be able to access information through digital television. Being younger this type is more likely to be living at home with parents than in a shared rented flat.

Group D : E for entertainment and shopping

This group includes a number of moderately well paid blue collar workers for whom the Internet and personal computing provide important leisure activities. This group tends to use the Internet not for obtaining information about products or for learning, but rather to provide access to music, games and general entertainment. People in this group are smart enough to learn new methods of accessing what they want but they are not necessarily interested in technology for its own sake. Besides providing a form of personal relaxation they also see the computer as a resource for family entertainment.

Members of this group are found among areas of cheaper owner occupied housing, particularly in neighbourhoods with high proportions of households with children.

Type D12 : Small time net shoppers
This type comprises many younger and middle aged men who particularly rely upon the Internet to buy music, books and videos. They are also active Internet purchasers of computer games and of fashion wear. This group is happy to undertake a wide variety of transactions on the Internet but tends not to be professionally involved in the development of information technology when at work

Type D13 : E for entertainment
Members of this type are not currently particularly active users of electronic technologies but are very interested in considering the purchase of new or enhanced products, from the range of mobile and personal computer devices. Many access the Internet using broadband and a high proportion purchase computer games. However this type is less interested in using the Internet for shopping, seeing it primarily as a leisure and entertainment medium.

Group E : E-independents

This group tends to take a rational and considered view of electronic communications and technologies. These people are not interested in mobile phones, texting or the Internet as lifestyle accessories; they do not feature as major topics of conversation within the social networks to which they belong and they do not provide a significant focus for leisure activity. However people are reasonably well equipped and use the Internet to search for information, to buy products and to undertake transactions where there are obvious efficiency benefits.

Type E14 : Rational utilitarians
This type tends to have access to the Internet at home and to use it extensively for shopping for groceries, wines, apparel, books and holidays, and for transacting financial services. Many of these people live in the countryside and

beyond the reach of cable television services. These people do not tend to use computers for playing games or as a form of leisure activity. Not being particularly heavy readers of computer magazines, these people treat the computer as a tool rather than as an end in itself.

Type E15 : Committed learners
This type consists of well educated, urban professionals with a high proportion of middle aged females, who use the Internet both for ordering and for information. Many of them have access to email and the Internet at work and consider information technology as a natural method of acquiring information – both as consumers and as emerging professionals. They tend to have access to technology that they are comfortable with and are less concerned than other groups about peer group opinion or the outward visible features of electronic devices.

Type E16 : Light users
This type contains many people who have access to electronic technologies but who are not very heavy users of them. Mostly in late middle age, these people do not view technology as a leisure activity and are not influenced by fashions or the need to keep up with peer groups. This type, though it does have access to the internet, tends not to use it to purchase games, fashion wear, videos or holidays, preferring to deal with organisations directly. However the type does purchase flowers over the Internet.

Group F : Instrumental E-users

This group tends to use electronic technologies for purely instrumental purposes, because they provide a practical method of saving time or money. They have plenty of other leisure activities that they enjoy and tend to be light television watchers. However they find the Internet useful for purchasing on line and they are smart enough to realise that they can drive better deals when purchasing goods and services if they fore-arm themselves with consumer information. Generally they use the net to undertake transactions and manage their personal finances rather than to explore.

This group contains mostly people in well off, middle class, owner occupied suburbia. Many have children.

Type F17 : Computer magazine readers
This type contains mostly middle aged users of electronic technology. They are people who have access to personal computers and the Internet and are interested in the features and functions of technologies. Many members of this type read magazines and purchase additional software and hardware over the Internet, but they are more oriented to the use of the Internet for personal finance

transactions than for purchasing. For example this type does not purchase children's wear or apparel over the Internet and is a low user of on line grocery shopping services. This type is a good market for Internet banking services.

Type F18 : E for financial management
This type contains mostly young people who work in companies which provide them with access to mobile phones, email and Internet access. It seems that many of these people lead lives which involve substantial amounts of travel between locations. Although competent in the use of electronic technologies they are not heavy purchasers of products through the Internet. However they are very heavy users of on line financial services. Flexibility is an important value for this type who feel the need to keep in constant touch with providers of information relevant to their daily lives.

Type F19 : On-line apparel purchasers
This type consists of well educated young professionals, many of them women, who are confident users of electronic technologies and communications. They use the Internet for purchases across a wide range of product categories, but in particular for children's products and fashion wear. They tend not to use this medium to purchase wines or insurance. Many members of this type look after children at home and do not have access to electronic technologies at work. They are not particularly interested in computer magazines.

Type F20 : E-exploring for fun
This type really enjoys the use of the computer to purchase products and services, making very high levels of on-line purchasing in virtually every product category – including traditional male purchases such as wines and insurance, computer games, videos and software, and traditional female purchases such as apparel and children's products. This type also likes to use the computer for personal banking services, but is not especially likely to be interested in cable television or mobile telephony. The majority are men, many of whom are in their thirties.

Group G : E-business users

This group includes many people who use electronic technologies in order to run their business. These may be people working in a technology related business or in a small business which needs to keep in electronic contact with its suppliers or its customers. Many of this group are self employed and make relatively little use of the technology as a leisure activity.

The group is well represented in upper income neighbourhoods attracting older professionals as well as in the countryside.

Type G21: Electronic orderers
This type is very likely to have a computer connection at home, but is likely to

make only light use of it. Few members of this group have access to email and the Internet at work but not at home. Many of this type own small businesses, and work and live outside London. Many are also farmers or proprietors of small establishments, who use technology to manage the administration of their businesses. The majority are male.

Group H : E- experts

Members of this group have every confidence in their abilities to undertake on-line transactions and to make full use of electronic technologies. These are the types of people who are able to make use of personalisation and configuration options. They enjoy exploring the features in electronic menus and will navigate them in an efficient manner. They prefer on line to inter-personal sources of information and make use of the Internet as an information source for obtaining best value for money. These people are heavy email users. Many of them are involved in the development of information technology applications at work, and see leisure time spent on electronic technologies as enhancing their human capital. Many recent graduates belong to this group.

This group is particularly concentrated in large cities and in the South East of England.

Type H22 : E-committed

This type finds it easy to acquire and master new technologies. The use of electronic technologies fits comfortably with the lifestyle which these people enjoy, which has a modern edge to it. These people rely on the Internet for information, though to a slightly lesser extent than those in type H23, and are active purchasers of goods and services over the Internet. Many of these people live in rented flats or are first time buyers on modern estates, have mortgages and children and feel the need to be familiar with information technology in order to advance their careers.

Type H23 : E - professionals

This type views the Internet and associated technologies as a indispensable basis of living. They use the Internet and new technologies in their professional lives, are constantly transferring numeric data as well as text messages, and are confident electronic orderers of specialist merchandise such as books and music. They are mostly young people, in and out of the office, who know how to access emails from locations other than their work and home. Young and well educated, a high proportion are students and single graduates, many of whom work in the new professions.

Table 3: Population sizes and gender mix of the 23 clusters

E-Cluster	Code	Weighted Population Size	% Pop.	% Male	% Female	% Unknown
Low technologists	**A 01**	2,544,078	7.19	28.9	70.4	0.69
Cable suffices	**A 02**	2,857,118	8.08	52.0	47.5	0.50
Technology as fantasy	**A 03**	1,916,191	5.42	44.2	54.8	0.97
Mobile's the limit	**A 04**	3,639,313	10.29	43.8	55.8	0.46
Too old to be bothered	**A 05**	780,343	2.21	42.7	56.6	0.69
Elderly marginalised	**A 06**	1,815,022	5.13	45.1	53.1	1.81
The net : what's that?	**B 07**	817,090	2.31	47.4	50.5	2.07
Mobile explorers	**B 08**	1,227,344	3.47	19.1	80.4	0.50
Cable tv heartland	**B 09**	1,037,479	2.93	43.0	53.2	3.83
E-bookers and communicators	**C 10**	1,052,948	2.98	39.5	59.0	1.42
Peer group adopters	**C 11**	874,385	2.47	45.7	52.6	1.68
Small time net shoppers	**D 12**	2,848,868	8.05	45.6	52.9	1.55
E for entertainment	**D 13**	2,097,600	5.93	56.6	42.5	0.92
Rational utilitarians	**E 14**	1,344,394	3.80	48.3	50.2	1.48
Committed learners	**E 15**	1,336,877	3.78	28.9	70.6	0.47
Light users	**E 16**	2,268,702	6.41	45.4	54.0	0.63
Computer magazine readers	**F 17**	992,291	2.81	58.9	40.3	0.85
E for financial management	**F 18**	375,249	1.06	57.5	41.9	0.64
On line apparel purchasers	**F 19**	1,633,201	4.62	16.1	83.4	0.55
Exploring for fun	**F 20**	869,252	2.46	61.3	37.7	0.98
Electronic orderers	**G 21**	1,937,974	5.48	77.4	21.8	0.82
E - committed	**H 22**	925,496	2.62	46.3	52.2	1.45
E - professionals	**H 23**	178,991	0.51	42.6	54.1	3.31

Table 4: Age mix of the 23 clusters (base score 100: thus a value of 50 reveals only half as many occurrences in a given age group as expected given the national demographic profile, while a value of 200 reveals twice as many occurrences as expected).

E-Cluster	Code	18-25	26-35	36-45	46-55	56-65	66+
Low technologists	**A 01**	54	43	37	40	66	246
Cable suffices	**A 02**	21	17	16	92	315	86
Technology as fantasy	**A 03**	26	26	28	45	190	184
Mobile's the limit	**A 04**	39	25	23	24	38	299
Too old to be bothered	**A 05**	8	7	8	14	83	304
Elderly marginalised	**A 06**	52	26	24	35	177	194
The net : what's that?	**B 07**	21	15	33	357	174	19
Mobile explorers	**B 08**	175	216	246	34	19	8
Cable TV heartland	**B 09**	292	177	149	65	39	27
E-bookers and communicators	**C 10**	223	258	169	34	20	13
Peer group adopters	**C 11**	263	261	135	41	22	16
Small time net shoppers	**D 12**	206	176	190	68	38	26
E for entertainment	**D 13**	190	156	206	94	33	22
Rational utilitarians	**E 14**	52	55	87	136	190	69
Committed learners	**E 15**	33	14	124	352	104	11
Light users	**E 16**	12	12	14	201	265	64
Computer magazine readers	**F 17**	55	25	137	297	104	21
E for financial management	**F 18**	50	141	197	176	67	7
On line apparel purchasers	**F 19**	167	269	185	34	20	13
Exploring for fun	**F 20**	89	171	237	114	31	12
Electronic orderers	**G 21**	125	231	224	44	25	16
E - committed	**H 22**	298	282	104	37	22	13
E - professionals	**H 23**	377	239	92	53	26	13

Table 5: Some illustrative characteristics of Groups 1 and 8 of the classification (index values calculated as in Table 4).

Behaviour	**Group 1**	**Group 8**
YOUR HOME - EMAIL AT WORK	30	519
YOUR HOME - EMAIL AT HOME	54	200
MONEY & INVESTMENT - FREQUENTLY USE INTERNET/PC BANKING	37	288
YOUR HOME – 2-5 HOURS SPENT PER WEEK ON INTERNET	54	163
HOME SHOPPING - BOUGHT GOODS/SERVICES 2-3 TIMES USING INTERNET	41	283

Figure 1: Cumulative population falling into pixel categories, ordered from the largest to the smallest. The plot is truncated at the pixel ranked 3,700

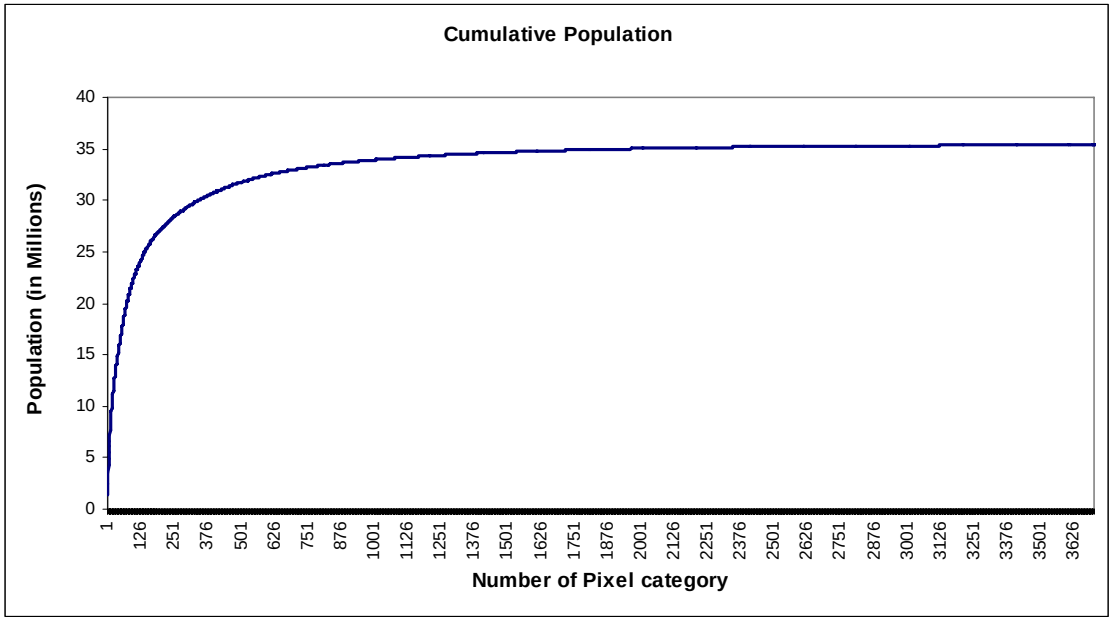


Figure 2: Tree profile of the E-Society classification

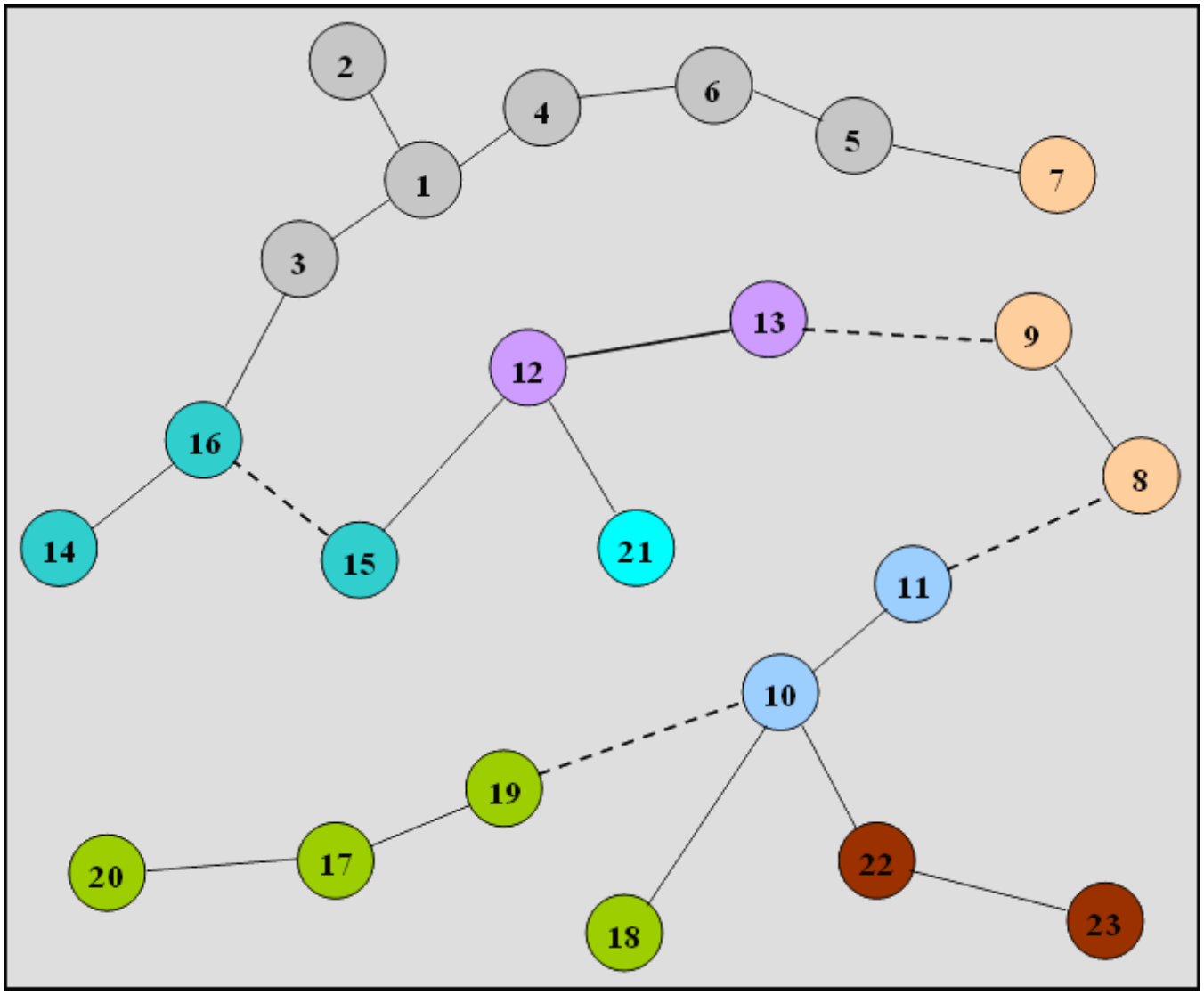


Figure 3: The Great Britain geography of two groups in the classification: (A) the E-unengaged and (B) E-experts.

(A)

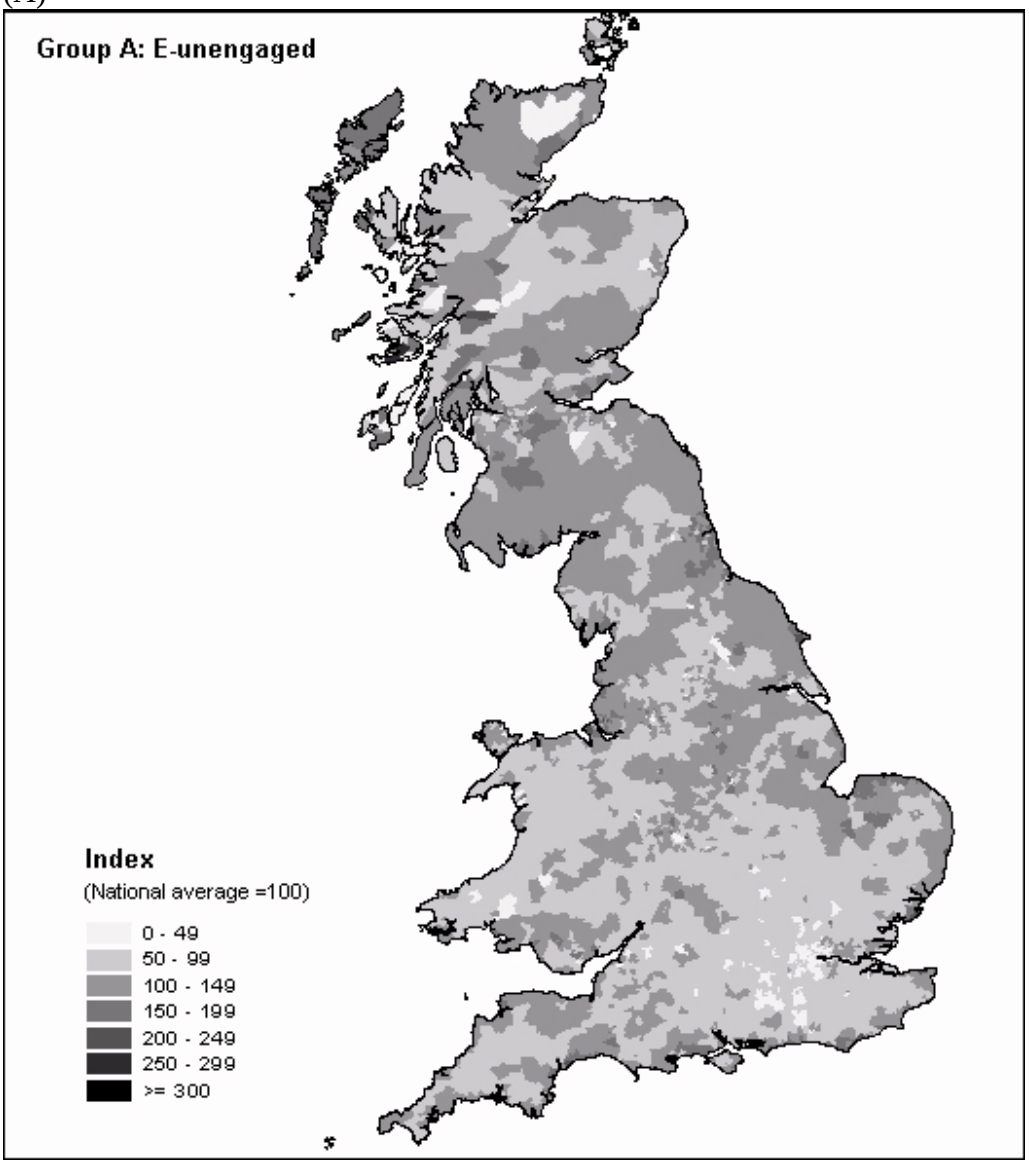


Figure 3 (continued)
(B)

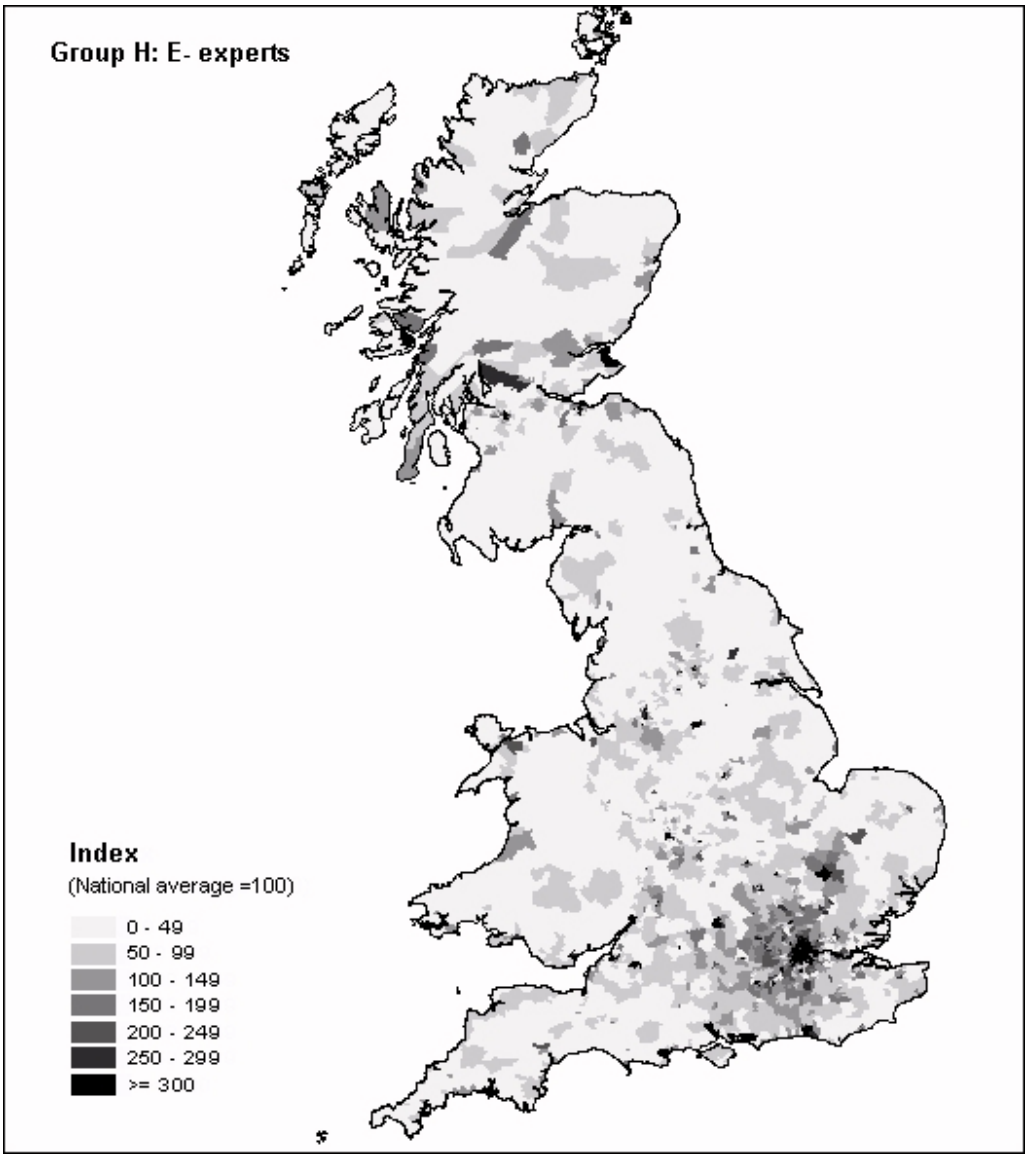
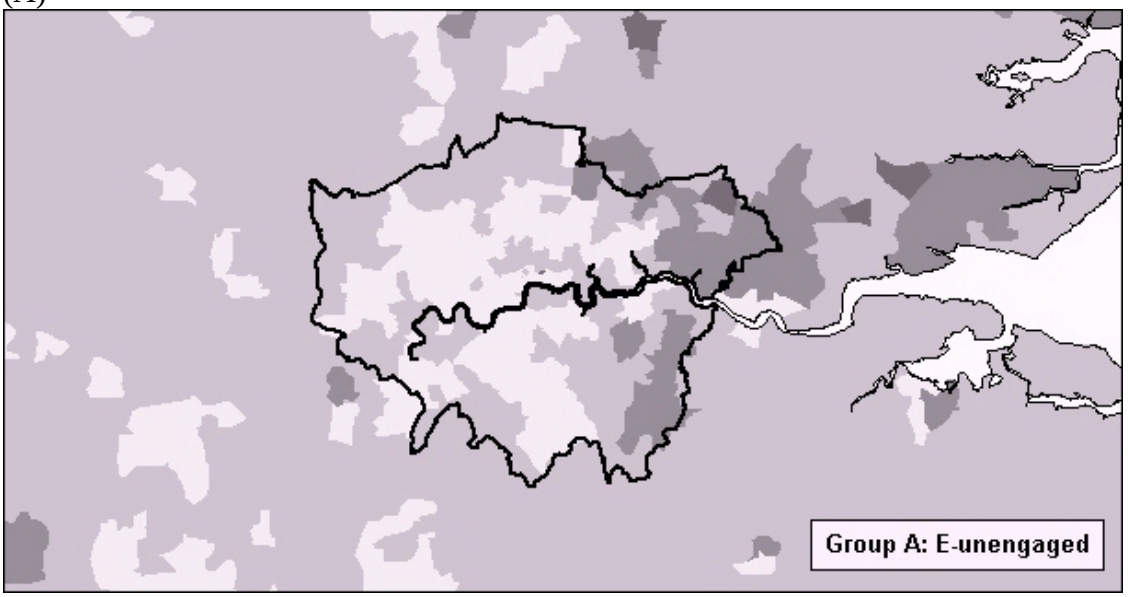


Figure 4: The regional geography of two groups in the classification: (A) the E-unengaged and (B) E-experts.

(A)



(B)